CERES Data Management Working Group Status Report

CERES Science Team Meeting

October 29, 2013

Chris Harris

Data Management Team

Systems/ Optimization:
Nelson Hillyer
Josh Wilkins

Production / Optimization:
Lisa Coleman
Carla Grune

Instrument:
Denise Cooper
Thomas Grepiotis
Dianne Snyder

Convolution:
Igor Antropov

Product Web Support:
Churngwei Chu
Ed Kizer
Cristian Mitrescu

Configuration Management:
Tammy Ayers
Joanne Saunders

ERBElike:
Dale Walikainen
Jeremy Lande

Inversion:
Victor Sothcott

SARB:
Tom Caldwell

Task Manager:
Walt Miller

Clouds:
Sunny Sun-Mack
Ricky Brown
Yan Chen
Liz Heckert
Rita Smith
Sharon Gibson

TISA Gridding:
Raja Raju
Forest Wren

TISA Averaging:
Cathy Nguyen
Dennis Keyes

Overview

- Software Status Update
- Current Product Availability
- Recent DMT Activity
- Production Processing Automation
- Edition 3 & Edition 4 Planned Milestones
- NPP Planned Milestones

CERES Production Software Overview

Subsystem Number	Subsystem Name	Number of PGEs	LOC (to nearest 1K)	Publicly Available Data Products	Product Frequency
1	Instrument/Pre-Processor	1	9K		1/day
1	Instrument	5 + lib	164K	BDS	1/day
2	ERBE-like/ Inversion	5 + lib	31K	ES-8	1/day
3	ERBE-like/ TSA	2	14K	ES-9, ES-4	1/month
4.1 – 4.4	Clouds/VIIRS Subset Code	1	21K		12/hour
4.1 – 4.4	Clouds	11	358K		1/month
4.5 – 4.6	Inversion	11	110K	SSF	1/hour
5	SARB	1	36K		1/hour
6 & 9	TISA-Gridding	7	34K	SSF1deg-Hour, ISCCP-D2like-Day/Nit	60/month, 36/month, 1/month
11	GGEO	1	6K	ISCCP-D2like-GEO	1/month
7.2	Synoptic SARB	1	47K		1/month
7.1 & 8 10	TISA-Averaging	4	102K	SSF1deg-Day, SSF1deg-Month, SYN1deg-(3Hour, M3hour, Mhour, Month)	1/day, 1/ month, 1/ month 5/month
12	MOA	2	14K		4/day
	CERESlib		121K		N/A
	Total	**53**	**1,046K**		

System Migration

- Migration from Magneto P4 system to AMI-P P6/x86 system nearly complete
- Edition 1-CV Clouds and Inversion subsystem code only code remaining active on Magneto P4
  - Inversion code delivered to AMI-P November 2012
  - Clouds code delivery expected October 2013
- Magneto transition expected fully complete February 2014
    - Power Off Ready (“CERES is not last man standing”)
  - New processors for ASDC production coming (2014)
    - IBM Power 7+ processors (128 cores)
    - Intel Sandy Bridge (E5-2690) processors (64 cores)
    - Porting and testing at SUSE 11 SP2 required

Current Data Availability

Product	Platform	Processed through	Publically Available
BDS (Ed 3)	Terra Aqua	Apr 30, 2013 Apr 30, 2013	Yes
SSF (Ed 3)	Terra Aqua	Jun 30, 2013 Jun 30, 2013	Yes
SFC (Ed 3)	Terra & Aqua	Mar 2013	Yes
SSF1deg-Month (Ed 3)	Terra & Aqua	Mar 2013	No
SYN1deg (Ed 3)	Merged	Mar 2013	Yes
ISCCP-D2like Products (All)	Terra & Aqua GEO Mrg	Sep 2012 Jun 2012 Jun 2012	Yes
SSF Edition 4 Beta 2	Terra Aqua	Jan 31, 2005 Feb 28, 2003	No
SSF Edition 1-CV	Terra Aqua	Jun 30, 2013 Jun 30, 2013	No

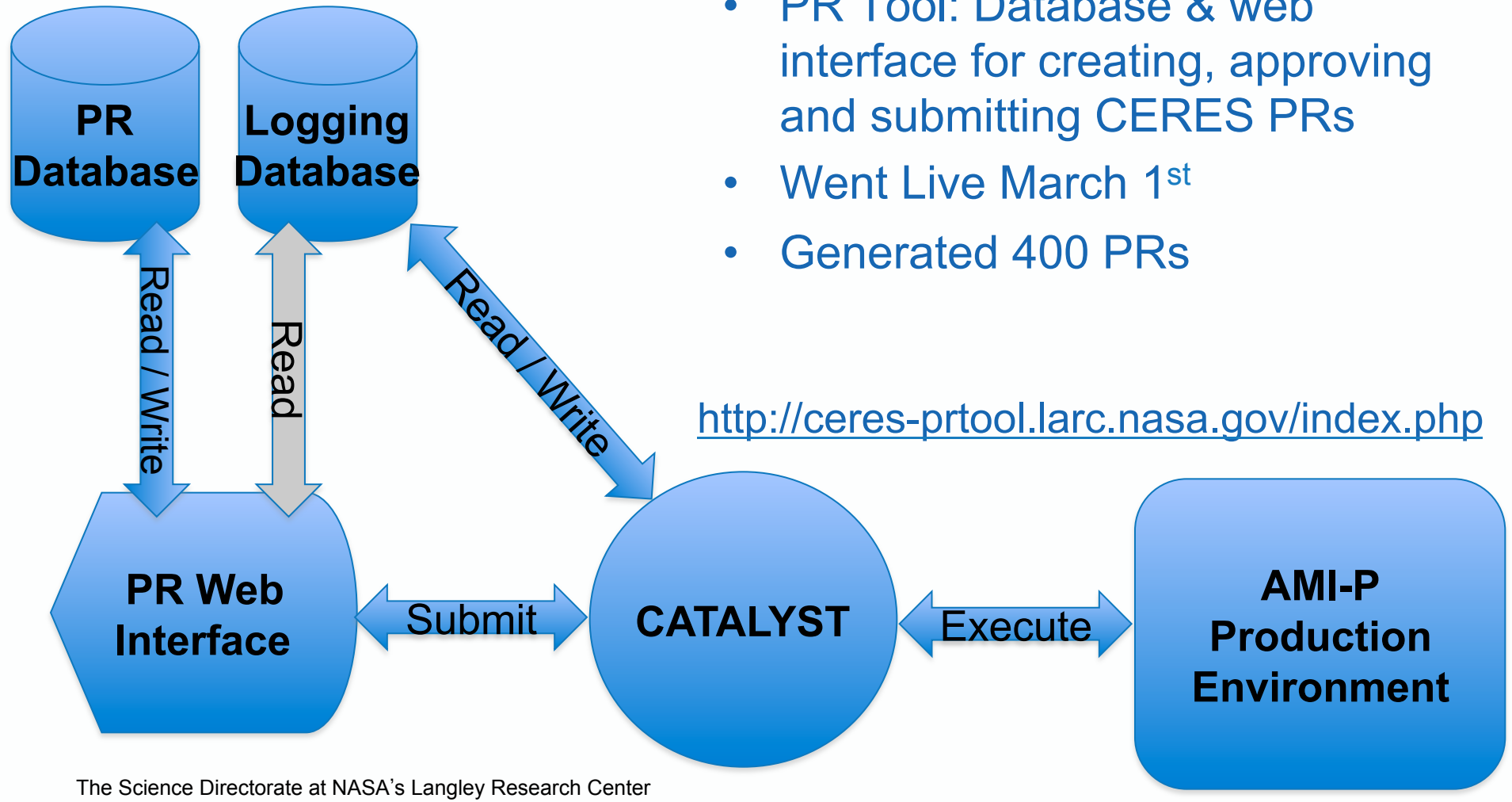
Recent Activity

16 Software & Data Deliveries Since 5/2/2013

- CATALYST & PR Tool (3)
  - PR DB New Clouds Startup Variables, Operator Comment Field, and Modified Clouds Main Processor Variables
- Instrument (1)
  - Variable gains for Ed1 NPP
- ERBE-like (1)
  - Command line option for Ed4
- Clouds (1)
  - P4 Migration
- Inversion (1)
  - Matching procedure change
- Tisa Gridding (1)
  - Cross track file update
- Tisa Averaging (2)
  - SORCE Update, Add new satellite ID
- SARB (4)
  - MATCH Files mandatory, MATCH file deliveries
- GGEO (1)
  - Ed3 Updates
- Perl_Lib (1)
  - CATALYST support

Production Automation Effort

Production Request (PR) Tool



- PR Tool: Database & web interface for creating, approving and submitting CERES PRs
- Went Live March 1st
- Generated 400 PRs

http://ceres-prtool.larc.nasa.gov/index.php

CATALYST

- **C**ERES **A**u**T**om**A**ted job **L**oading s**YST**em (CATALYST)
- Automation Framework for PGE Execution, Coordination & Logging that:
- Modularly implements science software dependency logic
- Leverages existing CERES production infrastructure
  - Ingest CERES PR to create collection of jobs
    - Build all possible individual jobs
    - Identify job input dependencies
  - Execute on AMI-P cluster
    - Send jobs to Sun Grid Engine
    - Utilize existing job submission scripts
  - Initiate ANGe ingest wrapper scripts
- Provides graphical interface for users to manage system

CATALYST (Cont'd)

- Automation Framework for PGE Execution, Coordination & Logging that:
- Dynamically manages job input dependencies and determine when jobs are runnable
  - CATALYST jobs wait for predecessor jobs to complete
    - Predecessor jobs can be:
      - Internal to CATALYST
      - External to CATALYST (via a backlogging interface)
    - Jobs broadcast completion status to follow-on jobs for rapid follow-on execution
- Stores job execution state long-term in a job logging database

CATALYST Implementation

- Primary languages Perl & Java
  - Perl ~13.8K LOC
  - Java ~6.9K LOC
- Defines an Application Programming Interface (API) using externally accessible XML-RPC
  - Allows users to inspect and modify job execution status programmatically in any language (with XML-RPC libraries)
  - Enables development of modular tools to function in concert with CATALYST – backlogging component
- Uses existing AMI LDAP service for user authentication
  - Permissions defined in access control list

CATALYST Build 1 Status

- Build 1 delivered for testing 2/25
- Internal CM testing and rework 2/26 – 3/21
- Test Readiness Review 3/22
- ASDC Operational Testing 3/27 – current
- Four software patches (11/18 for fourth)
- Go-Live – 1/22

- Concurrent month Terra and Aqua SSFs (8 PGEs Clouds and Inversion) produced in 30 hours for all 4 instruments.

CATALYST Builds

- Build 1.1: Functionality
  - API to update Logging DB with “known missing” cases
  - Job submission balancing function
  - Expected Live April 2014
- Build 1.2: PGEs Only
  - Edition 4 Inversion PGEs
  - Expected Live July 2014
- Build 1.3
  - Edition 1-CV processing stream
    - Instrument, ERBElike, Clouds, Inversion, MOA
  - Expected Live October 2014

Terra / Aqua Edition 3 & 4 Planned Milestones

Edition 3 & 4 Planned Milestones

Product	Science Delivery to DMT	Target Public Release
Ed3 Flux-By-Cloud-Type	Oct 25, 2013	Dec 20, 2013
Ed4 Spec. Response func	Nov 1, 2013	N/A
Ed4 Inversion	Nov 22, 2013 (ADM & SOFA groups)	Feb 7, 2014
Ed4 SSF1deg-Hour	Feb 28, 2014	May 23, 2014
Ed4 SSF1deg-Month	Mar 31, 2014	Jun 27, 2014
Ed4 TSI	Feb 28, 2014	Aug 20, 2014
Ed4 SYNI	Apr 11, 2014	
Ed4 SYN1deg	Apr 25, 2014	
Ed4 ISCCP-D2like (all 3)	May 30, 2014	Aug 22, 2014
Ed4 CRS	Jul 31, 2014	Dec 22, 2014
Ed4 Flux-By-Cloud-Type	Sep 30, 2014	Nov 28, 2014

NPP Planned Milestones

NPP Planned Milestones

Product	Science Delivery to DMT	Target Public Release
Ed1 Clouds (SSF)	Nov 22, 2013	May 16, 2014
Ed1 SOFA Code (SSF)	Feb 10, 2014	
Ed1 Gains (BDS)	Nov 29, 2013	Dec 18, 2013
Ed1 SSF1deg-Hour	March 14, 2014	May 27, 2014
Ed1 CRS	March 24, 2014	May 30, 2014
Ed1 SSF1deg-Month	April 11, 2014	June 13, 2014
Ed1 SYNI	April 30, 2014	July 7, 2014
Ed1 TSI & SYN1deg	May 9, 2014	

Rotational Assignments

Rotational Assignment Started May 13th

•Chris Harris will join the CERES team as Data Management Working Group Lead

•Jonathan Gleason will join the ASDC team

The goal is to provide a broader range of experience for each individual, while giving them insight into each other's team processes.